

## Aberystwyth University

### *Breast Ultrasound Region of Interest Detection and Lesion Localisation*

Hoon Yap, Moi; Goyal, Manu; Osman, Fatima; Marti, Robert; Denton, Erika R. E.; Juetten, Arne; Zwiggelaar, Reyer

*Published in:*  
Artificial Intelligence in Medicine

*DOI:*  
[10.1016/j.artmed.2020.101880](https://doi.org/10.1016/j.artmed.2020.101880)

*Publication date:*  
2020

*Citation for published version (APA):*

Hoon Yap, M., Goyal, M., Osman, F., Marti, R., Denton, E. R. E., Juetten, A., & Zwiggelaar, R. (2020). Breast Ultrasound Region of Interest Detection and Lesion Localisation. *Artificial Intelligence in Medicine*, 107, [101880]. <https://doi.org/10.1016/j.artmed.2020.101880>

#### **Document License** CC BY-NC-ND

#### **General rights**

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400  
email: [is@aber.ac.uk](mailto:is@aber.ac.uk)

# Breast Ultrasound Region of Interest Detection and Lesion Localisation

Moi Hoon Yap<sup>a,\*</sup>, Manu Goyal<sup>a</sup>, Fatima Osman<sup>b</sup>, Robert Martí<sup>c</sup>, Erika Denton<sup>d</sup>, Arne Juetten<sup>d</sup>, Reyer Zwiggelaar<sup>e</sup>

<sup>a</sup>*Department of Computing and Mathematics, Manchester Metropolitan University, UK*

<sup>b</sup>*Department of Computer Science, Sudan University of Science and Technology, Sudan.*

<sup>c</sup>*Computer Vision and Robotics Institute, University of Girona, Spain*

<sup>d</sup>*Norfolk and Norwich University Hospital Foundation Trust, Norwich, UK*

<sup>e</sup>*Department of Computer Science, Aberystwyth University, UK.*

---

## Abstract

In current breast ultrasound Computer Aided Diagnosis systems, the radiologist preselects a region of interest (ROI) as an input for computerized breast ultrasound image analysis. This task is time consuming and there is inconsistency among human experts. Researchers attempting to automate the process of obtaining the ROIs have been relying on image processing and conventional machine learning methods. We propose the use of a deep learning method for breast ultrasound ROI detection and lesion localisation. We use the most accurate object detection deep learning framework – Faster-RCNN with Inception-ResNet-v2 – as our deep learning network. Due to the lack of datasets, we use transfer learning and propose a new 3-channel artificial RGB method to improve the overall performance. We evaluate and compare the performance of our proposed methods on two datasets (namely, Dataset A and Dataset B), i.e. within individual datasets and composite dataset. We report the lesion detection results with two types of analysis: 1) *detected point* (centre of the segmented region or the detected bounding box) and 2) Intersection over Union (*IoU*). Our results demonstrate that the proposed methods achieved comparable results on *detected point* but with notable improvement on *IoU*. In addition, our proposed 3-channel artificial RGB method improves the *recall* of Dataset A. Finally, we outline some

---

\*Corresponding author

*Email address:* `m.yap@mmu.ac.uk` (Moi Hoon Yap )

future directions for the research.

*Keywords:*

Breast ultrasound, breast cancer, object detection, region of interests

---

## 1. Introduction

Breast cancer is a common disease for women and is considered to be the second leading cause of death worldwide [1]. According to Breast Cancer Now [2], breast cancer is the most common cancer in the UK. Ultrasound is the complementary modality to the standard imaging method (two view mammography) in breast cancer diagnosis [3, 4]. It is the most widely used in clinical practice [5] compared to other alternatives such as tomosynthesis and magnetic resonance imaging. Due to the fact that early detection plays a main role in avoiding breast cancer deaths and increases the proportion of healing and recovery, there has been increasing interest in using ultrasound to aid in the early detection of breast cancers over the past few years [6, 7].

In Breast Ultrasound (BUS), radiologists are trained in interpreting the sonographic features [8]. In current practice, the clinician scans the breast and takes static images. The radiologist will assess and annotate the BUS images. Computer Aided Diagnosis (CAD) systems are then can be used as a “second reader” for computerized medical imaging analysis [9]. These systems are based on the assumption that the radiologist detects an abnormality and preselects a region of interest (ROI). Figure 1 shows BUS images with manual pre-selected ROIs marked with ‘+’ and ‘x’.

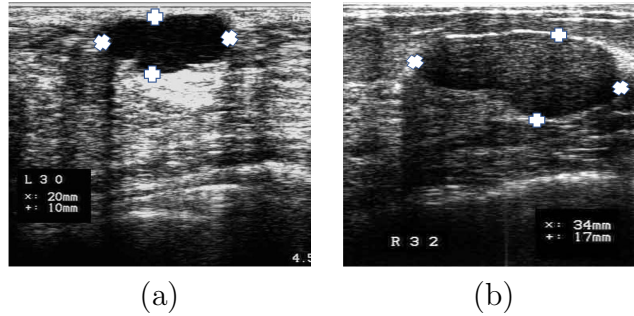


Figure 1: Examples of BUS images with manual pre-selected ROIs marked with ‘+’ for the upper and lower points for the lesion, and ‘x’ for the leftmost and rightmost points of the lesion. Please note that the annotations were embossed for better visualisation.

Previous work attempted to automate the process of ROIs selection [10, 11, 12, 13]. These methods were based on multi-stage image processing and/or machine learning approaches. Deep learning has gained popularity in biomedical image analysis and has achieved good results in classification [6, 14] and BUS semantic segmentation [15]. Yap et al. [7] compared the performance of lesions detection algorithms and showed that deep learning approaches are more accurate and robust across datasets. However, the limitations of their work were: 1) they detected the lesions by using segmentation approaches but not an object detection approach; and 2) they evaluated the performance based on *detected point* (centre of the segmented region) [7], not the overlap of the regions.

According to state-of-the-art BUS lesion detection [6, 16], a ROI is defined as a bounding box circumscribing the lesion. This paper focuses on the automatic detection of such ROIs. We propose the use of the Faster-RCNN Inception-ResNet-v2 approach [17] for BUS lesion detection. The key contributions are:

1. We automate the ROI detection using a popular deep learning approach, this is the first attempt in automation of BUS ROI detection using Faster-RCNN Inception-ResNet-v2.
2. We propose two approaches to overcome the issue of lack of BUS data. First we apply a transfer learning approach and then we propose a new 3-channel artificial RGB method to improve the quality of results.
3. We evaluate and compare the performance of our proposed method on two datasets - within individual datasets and composite dataset. As existing approaches do not focus on ROI bounding box detection, we compare the performance of our proposed methods with FCN-AlexNet.

## 2. Related Work

In current practice, the clinical expert manually locates rectangular sub-images [18, 19] to locate ROIs on BUS images. However, in large-scale studies, this step is time-consuming. Hence, researchers [20, 21, 10] have developed algorithms to locate the ROIs automatically. Within fully automated ROI detection, there are two types of ROI: 1) ROI as an initial contour of the lesion [20, 21, 22, 23]; and 2) ROI as a rectangle region containing both lesion and some background information [10, 12]. In this section, we review research on both ROI definitions.

Table 1: A Comparison of Dataset A and Dataset B.

Comparison	Dataset A	Dataset B
Capture Devices	B&K Medical Panther 2002 and B&K Medical Hawk 2012	Siemens ACUSON Sequoia C512 system
Transducer	8-12 MHz linear array transducer	8.5 MHz 17L5 HD linear array transducer
Year	2001	2012
Number of Images	306	163
Image size	$377 \times 396$	$760 \times 570$

55 In 1998, based on a single feature called the radial gradient index (RGI),  
 56 Kupinski et al. [20] developed a novel lesion segmentation technique. Us-  
 57 ing gray-level information, and prior knowledge of the shape of typical mass  
 58 lesions, a series of image partitions were created and the partition that max-  
 59 imised the RGI was selected. The method was tested on a database of  
 60 biopsy-proven, malignant lesions. According to their results [20], the RGI  
 61 segmentation algorithm correctly segmented 92% of the lesions. Although  
 62 the work of Kupinski et al. [20] assessed the RGI filter in mammograms, it  
 63 was applied to BUS images in 2002 by Drukker et al. [21], where the use  
 64 of RGI filtering technique was investigated for automated lesion detection in  
 65 BUS. Using a database of 757 images from 400 patients, lesion candidates  
 66 were segmented from the background by maximising an average radial gra-  
 67 dient index for regions grown from the *detected point*. Initial RGI filtering  
 68 achieved a sensitivity of 87% at 0.76 false-positive detections [21].

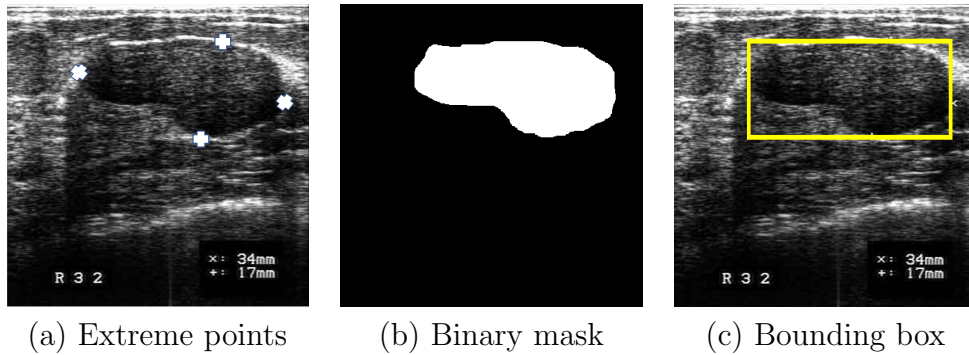


Figure 2: The ground truth format conversion of BUS datasets: (a) original extreme points; (a) original segmentation ground truth in binary mask form provided by Yap et al. [7]; and (c) conversion to bounding box as the ground truth for ROI detection and localisation.

In 2008, Yap et al. [10] proposed a novel approach for boundary detection of ROI in BUS images. In the preprocessing step, histogram equalization was applied, followed by a combination of nonlinear diffusion and linear filtering. Further to this hybrid filtering stage, the visually distinct areas of the BUS image were analysed using multifractals. In the final stage, region growing based segmentation was applied to partition the filtered BUS image using different threshold values. According to the assumption of Kupinski et al. [20], selection of the lesion was made by choosing the partition with the highest RGI. The work indicated that multifractal analysis could be useful for enhancing boundary detection in ultrasound images.

For the detection of masses, Ikedo et al. [24] used a feature based on the edge directions in each slice, and a method for subtracting between slices. In order to detect edges, a Canny edge detector was applied and morphology was used to classify the detected edges into two groups: near-vertical edges or near-horizontal edges. Subsequently, the near-vertical edges were used as cues, then using the segmented and the low-density regions, they were able to segment the located positions by a watershed algorithm, and mass candidate regions were detected. Finally, for the distribution between masses and false positives (*FPs*), rule-based schemes and a quadratic discriminant analysis were applied in order to remove *FPs*. Aiming to improve the screening performance and efficiency, the proposed scheme achieved sensitivity of 80.6% with 3.8 *FPs* per breast image.

A fully automated segmentation method was proposed in 2012 by Shan et al. [12]. Two main findings were introduced: an efficient ROI generation method and new features to characterise lesion boundaries were proposed. In order to develop an automatic ROI generation method, two steps were used, the first step was the automatic seed point selection and the second was a region growing step. Region growing was considered to be fast and simple, although its accuracy was not high, it was serving the purpose as it roughly located the lesion rather than finding the accurate boundary of it. Further, they combined traditional intensity-and-texture features and two proposed lesion features (phase in max-energy orientation and radial distance) were used to detect lesions by a trained artificial neural network. On a database of 120 images, the method improved the true positive (*TP*) rate from 84.9% to 92.8%, the similarity rate from 79.0% to 83.1% and reduced the *FP* rate from 14.1% to 12.0%.

In order to detect lesions in breast US images, with no need for any kind of human interaction or supervision, Pons et al. [25] proposed a feasibility

study by adapting a generic object detection technique, called Deformable Part Models (DPM). They provided an assessment of this methodology to lesion detection by applying it for the first time to US images, using a dataset of 100 images, all from different patients (50 were healthy tissue regions, 18 were malignant lesions, 32 were benign lesions). According to results for lesion detection, they showed the feasibility of their proposal and they achieved a sensitivity of 82% with 0.51 false-positive detections per image and an  $A_z$  value of 0.96.

Although research to date has demonstrated the feasibility to automate the ROI detection by using computer algorithms, like in similar medical image analysis research, there are some common issues:

1. Research was conducted within a single institution or hospital; code and datasets were not shared. Therefore, the research is not reproducible, and less straight forward to compare.
2. The use of performance metrics has not been consistent, i.e. some used  $FP$  rate, while others used  $FP$  per image; some reported sensitivity and specificity, while others used *recall* and *precision*.
3. The methods were mostly based on image processing and conventional machine learning. Although some researchers [7, 15, 6] have been working actively in deep learning for classification and segmentation, the use of deep learning for ROI detection in BUS is yet to be fully explored.

We address these issues by proposing the use of a popular deep learning method for ROI detection on two publicly available datasets, and we report the results with a variety of performance metrics. If the manuscript is accepted for publication, the codes will be made available on github.

### 3. Methodology

This section discusses the BUS datasets, the preparation of the ground truth labeling, the proposed ROI detection method (based on transfer learning, the 3-channel Artificial RGB image method and a Faster-RCNN approach) and the performance metrics for the ROI detection results.

#### 3.1. Datasets and Ground Truth

In general, ultrasound images are complex because of data composition, which can be described in terms of speckle information. Upon visual inspection, ultrasound images could be described as speckle noise that varies

141 between bright and dark degrees of grayscale. The two datasets (henceforth,  
 142 Dataset A and Dataset B) that we used in this paper were obtained from a  
 143 recent publication by Yap et al. [7]. They are referred to as Dataset A and  
 144 Dataset B and Table 1 compares the two datasets. The 306 images in Dataset  
 145 A are from 2001. Although Dataset A might not be a representative of clin-  
 146 ical practice, it is still interesting to test the robustness of machine learning  
 147 algorithms on different image resolutions. The 163 images in Dataset B are  
 148 from 2012 and have a higher image resolution. To standardize the image  
 149 resolution for our experiments, we have resized the images to  $500 \times 375$ . For  
 150 a detailed description and to download Dataset B, please refer to [7].

151 The ground truths provided in the BUS datasets are in the form of binary  
 152 masks of the lesions or with extreme points, as illustrated in Fig. 2(a). From  
 153 these extreme points, we generated rectangle bounding boxes around the  
 154 binary masks for ROI localisation. Fig. 2(b) illustrates an example of a  
 155 bounding box overlaid on the original BUS image. This is a mandatory step  
 156 as the bounding boxes are commonly used in computer vision as the ground  
 157 truth labels to train the object detection algorithms.

### 158 3.2. Transfer Learning

159 To obtain good performance, current state-of-the-art deep learning meth-  
 160 ods require large-scale datasets to train the model [26]. In natural images,  
 161 large-scale datasets exist such as ImageNet [27] and the MS-COCO dataset  
 162 [28]. ImageNet [27] consists of more than 1.5 million images for the clas-  
 163 sification of 1000 pre-defined classes [27] and the MS-COCO dataset [28]  
 164 consists of 328,000 images with 91 common object categories. To use these  
 165 pre-trained models for our proposed BUS ROI lesion detection framework,  
 166 we convert the original grayscale BUS images to 3-channel images ( $I$ ) by con-  
 167 catenating three single channel grayscale images ( $I_g$ ) from the BUS datasets,  
 168 as shown in Equation 1.

$$I = \text{Concat}(I_g, I_g, I_g) \quad (1)$$

169 where  $I$  is a 3-channel converted image from the concatenation of three orig-  
 170 inal grayscale images ( $I_g$ ).

171 Transfer learning is a popular technique in deep learning to overcome data  
 172 deficiency, where we can choose to transfer the features from a few convolu-  
 173 tional layers (partial transfer learning) or from all layers (full transfer learn-  
 174 ing) of a pre-trained model. For our proposed framework, we implemented



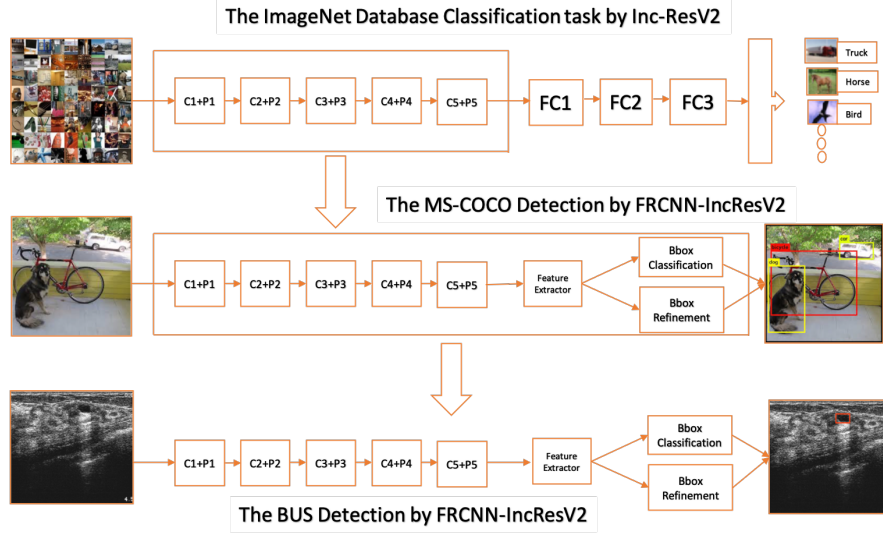


Figure 3: Overview of two-tier transfer learning used for ROI detection and localisation of BUS lesions.

two-tier transfer learning [29]. Firstly we used partial transfer learning by transferring the features only from the convolutional layers trained on the most significant classification challenge dataset - ImageNet. Then, we used full transfer learning from a model trained on MS-COCO object localisation dataset as shown in Fig. 3.

### 3.3. 3-channel Artificial RGB Image Method

In standard data augmentation techniques, the number of training images is increased with different image manipulation algorithms, including rotation, flipping and image filtering. Data augmentation has shown to be effective in improving the performance of deep learning algorithms. However, it has increased the time and memory requirements in training the algorithms. We propose a new 3-channel artificial RGB image method by concatenating the original image with two post-processed images. With this proposed technique, we maintain the number of training images, i.e. rather than concatenating the three grayscale images, we used two filtered images to concatenate with the grayscale image. The proposed 3-channel artificial RGB image ( $I_a$ ) is produced by concatenating a single channel grayscale image ( $I_g$ ), the sharpened image ( $I_s$ ) and the contrast enhanced image ( $I_c$ ), as shown in Equation 2.

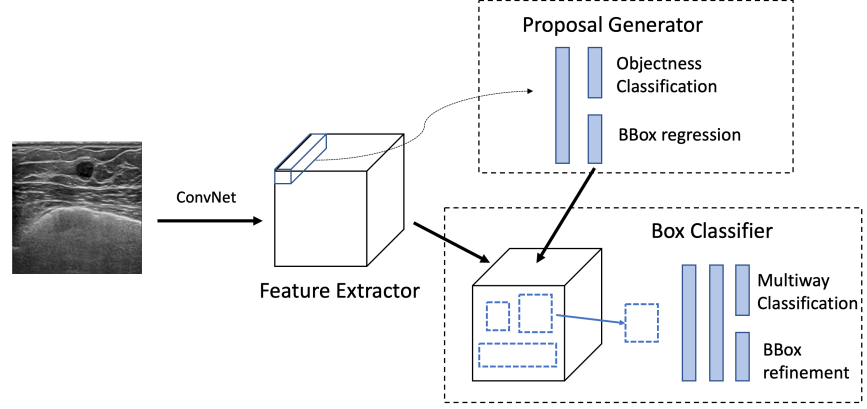


Figure 4: Overview of the proposed architecture (redrawn from [30]) for BUS experiments. The Proposal Generator generates Bounding Box (BBox) from the feature maps. The refinement and classification of BBox proposals are attained by Inception-ResNet-v2 to obtain the best accuracy of BBox.

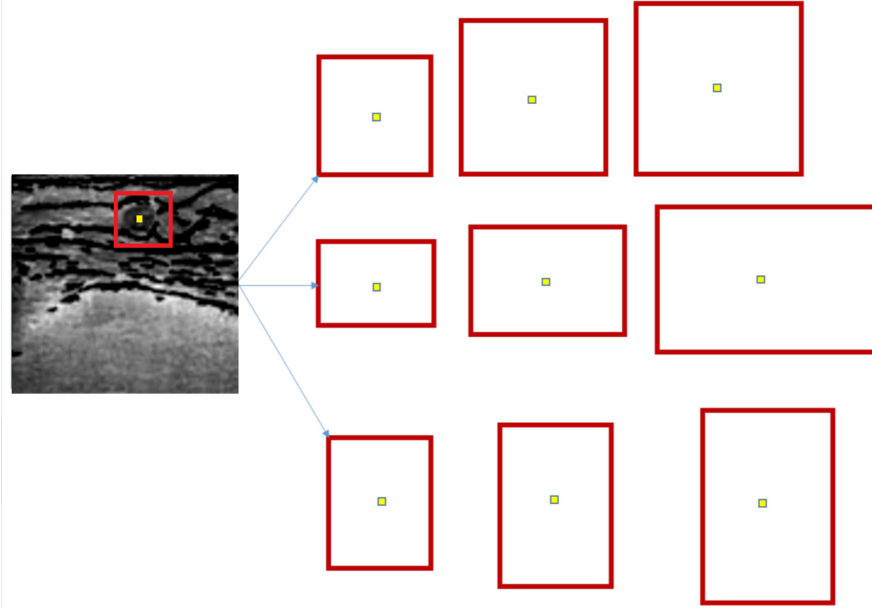


Figure 5: Nine different anchors are generated for a single point of the feature map.

$$I_a = \text{Concat}(I_g, I_s, I_c) \quad (2)$$

#### 194 3.4. *Faster-RCNN Inception-ResNet-v2 approach*

195 Faster-RCNN Inception-ResNet-v2 is one of the most accurate state-of-  
 196 the-art models for object localisation [30]. It has been successfully imple-  
 197 mented, e.g. in person detection [28] and diabetic foot ulcers localisation  
 198 [31]. In the earlier version of the Region Proposal Network (RPN), the first  
 199 step is to generate region proposals by selective search, then classify and de-  
 200 tect the object based on a Convolutional Neural Network (CNN) framework.  
 201 The core design of the Faster-RCNN was similar to the Region-based CNN,  
 202 i.e. hypothesise object regions based on the feature maps and then classify  
 203 them using the similar CNN. The benefit of Inception-ResNet-v2 [17] is it  
 204 combined the optimization benefits conferred by residual connections with  
 205 the computation efficiency of Inception units. Figure 4 illustrates the archi-  
 206 tecture for Faster-RCNN [32] with Inception-ResNet-v2 approach [17]. The  
 207 architecture of Faster-RCNN consists of three stages:

- 208 • First Stage: A pre-trained CNN (Inception-ResNet-v2) was used to  
 209 extract the convolutional feature map of BUS images from the last  
 210 convolutional layer for proposal generator (Second Stage) and BBox  
 211 classification and regression (Third Stage).
- 212 • Second Stage: The proposal generator is used to find a predefined  
 213 number of bounding box (BBox) proposals, may contain a lesion. An-  
 214 chors are fixed bounding boxes that are placed throughout the image  
 215 with different sizes (64px, 128px, 256px) and ratios (0.5, 1, 1.5) to find  
 216 lesions in the BUS image as shown in Fig. 5. Then, two layers (ob-  
 217 jectness classification layer and BBox regression layer) are used to find  
 218 the “objectness score” for these anchors to have a good set of BBox  
 219 proposals. For this stage, as BUS images have a very limited number  
 220 of lesions (mostly one lesion per image), we set the value of a number  
 221 of proposals to 100.
- 222 • Third Stage: Finally, these BBox proposals (from the Second Stage)  
 223 are then passed through a pre-trained CNN in the next step to extract  
 224 features for each proposal. The ROI pooling layer is used to produce  
 225 fixed-size feature maps from non-uniform inputs of proposals by per-  
 226 forming a max pooling operation. These features are finally used by the

Box Classifier (classification and BBox refinement layers) to refine and classify the proposals, which obtains the final accurate BBox regions. We only chose BBox regions with confidence equal to 90% or higher for final evaluation.

### 3.5. Performance Metrics

We used four popular performance metrics i.e. *Precision*, *Recall*, *F1-Score* and *False Positives per Image (FPI)* for the evaluation of BUS detection and localisation. The state-of-the-art BUS lesion detection research used *detected point* criterion [7]. However, the measurement based on the centre of detected bounding box or segmented region can be misleading. To overcome this issue, we use “overlap criterion” as an Intersection over Union (*IoU*) greater than 0.5 [33]. The *IoU* is defined by equation 3.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (3)$$

In the context of medical image analysis, *IoU* is known as the Jaccard Similarity Index or Jaccard Index. Based on *IoU* as the criteria, we calculate the following parameters:

1. *True Positives (TP)* defined as Bounding Boxes (BBox) that have *IoU* greater than 0.5 with the BB of the ground truth (GT).
2. *False Positives (FP)* defined as BBox that have *IoU* less than 0.5 with GT and also, the duplicate BB that have *IoU* with a GT that has already been detected.
3. *True Negatives (TN)*: In BUS datasets, all the images contain at least one lesion. This is due to current practice that the clinician will only save ultrasound images with a lesion. Hence, there were no normal images and we can not obtain *TN*.
4. *False Negatives (FN)* were calculated if there is no detection of the BBox produced by the algorithm.

The *Precision* was calculated by total number of correct BBox i.e. *TP* divided by the total number of ground truth i.e. *TP* and *FP*, as shown in equation 4. The *Recall* was the total number of correct detected bounding boxes (i.e. *TP*) divided by total number of detected bounding boxes (i.e. *TP*) and *FN*, as in equation 5. The last evaluation metric was the *F1-Score*, which was the harmonic average of *Precision* and *Recall* (see equation 6). The *F1-Score* is also known as Dice Coefficient Index in medical image analysis.

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1 - Score = \frac{2 \times (Recall \times Precision)}{Recall + Precision} \quad (6)$$

260 To compare with state-of-the-art methods, we also report our results as  
 261 in Yap et al. [7], i.e. detection is considered as a *TP* if the detection point  
 262 (centre of the detected bounding box) is placed within the ground truth  
 263 bounding box of an expert radiologist. Otherwise, it was considered to be  
 264 a *FP*. Figure 6 compares the differences between two criteria, where *IoU* is  
 265 more reliable in reporting the results.

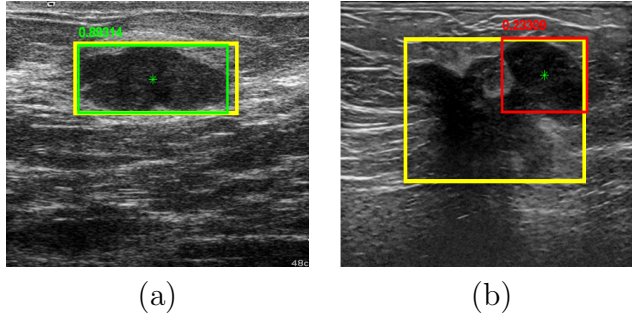


Figure 6: The yellow box indicates ground truth, the green ‘\*’ indicated the *detected point*, the green bounding box indicates true detection and the red bounding box indicates false detection: (a) this is an example of both *detected point* and *IoU* achieved agreement with a True Positive; and (b) this is an example where even the detected region is at the top right corner, the *detected point* calculated as true detection but the *IoU* has a more strict measurement and categorised it as a false detection.

### 266 3.6. Implementation

267 For consistency, we have evaluated all the methods using 5-fold cross-  
 268 validation on 3-channel grayscale datasets and 3-channel artificial RGB datasets.  
 269 For the composite dataset (combination of dataset A and dataset B), this  
 270 was not totally random as we needed to ensure the training set distribu-  
 271 tions consisted of both datasets. For the benchmark algorithm, we used the  
 272 Caffe framework [34] to implement the transfer learning FCN-AlexNet. We

repeated the experiment using similar settings as in [7], where the model was trained using stochastic gradient descent with a learning rate of 0.001, 60 epochs with a dropout rate of 33%. To convert the segmentation results produced by FCN-AlexNet, we used the similar method in converting the binary masks to ground truth bounding boxes, where the coordinates of the left most pixel, the top most pixel, the right most pixel and the bottom most pixel are used to form the bounding box (as illustrated in Fig. 2).

For the implementation of the Faster-RCNN Inception-ResNet-v2 approach (henceforth, FRCNN), we used the original parameters as in [17], with the learning rate of 0.001. We observed the models converged at 100 epochs. Our experiments were run on a GPU machine with the following configurations: (1) Hardware: CPU - Intel i76700@4.00 Ghz, GPU - NVIDIA TITAN X 12 GB, RAM - 32 GB DDR4 (2) Deep Learning framework: Tensor-flow.

## 4. Result and Discussion

We performed thorough evaluation within and between the datasets. We evaluated the results based on 5-fold cross validation on single datasets (solely on Dataset A and Dataset B ) and composite dataset (A+B). We reported the results of the individual dataset in the composite dataset experiment, which was (A+B) on A and (A+B) on B. We discuss the results in two detection methods, i.e. *detected point* and *IoU*. Then we perform visual comparison of the results.

### 4.1. Evaluation based on detected point

Table 2 shows the overall FRCNN results based on *detected point*. From the results of Yap et al. [7], the transfer learning FCN-AlexNet (henceforth, FCN-AlexNet) [35] outperformed Radial Gradient Index Filtering [21], Multifractal Filtering [10], Rule-based Region Ranking [12], Deformable Part Models [13], and two deep learning techniques (U-Net [36] and Patched-based LeNet [37]). To compare the performance of FRCNN on BUS lesion detection, we used FCN-AlexNet as the benchmark algorithm.

#### 4.1.1. Within dataset analysis

We observed all the methods were obtaining high *recall* and *precision* when evaluated based on *detected point*. Although the performance of FRCNN obtained the best results in this setting, the *recall* for FCN-AlexNet is comparable. Overall, FRCNN achieved the best *F1-Score* but FRCNN with

Table 2: Comparison of performance metrics based on *detected point* for ROI detection in BUS dataset. FRCNN is Faster-RCNN Inception-ResNet-v2 on concatenated grayscale BUS images whereas FRCNN (RGB) is Faster-RCNN Inception-ResNet-v2 on 3-channel artificial RGB BUS images. FCN-AlexNet represents transfer learning FCN-AlexNet. Bold indicates the best result for each category and underline indicates the best result for the Dataset.

Dataset	Method	<i>Recall</i>	<i>Precision</i>	<i>F1-Score</i>	<i>FPI</i>
A	FCN-AlexNet	0.9388	0.8365	0.8847	0.1961
	FRCNN	0.9236	<b><u>0.9408</u></b>	<b><u>0.9321</u></b>	<b><u>0.0621</u></b>
	FRCNN (RGB)	<b><u>0.9572</u></b>	0.9020	0.9288	0.1111
B	FCN-AlexNet	0.9080	0.8605	0.8836	0.1472
	FRCNN	<b>0.9141</b>	<b><u>0.9371</u></b>	<b><u>0.9255</u></b>	<b><u>0.0614</u></b>
	FRCNN (RGB)	0.8589	0.8861	0.8723	0.1104
(A+B) on A	FCN-AlexNet	0.9450	0.8351	0.8867	0.1994
	FRCNN	<b>0.9480</b>	<b>0.8857</b>	<b>0.9158</b>	<b>0.1307</b>
	FRCNN (RGB)	0.8746	0.8338	0.8537	0.1863
(A+B) on B	FCN-AlexNet	0.9325	0.7917	0.8563	0.2454
	FRCNN	<b><u>0.9632</u></b>	<b>0.8441</b>	<b>0.8997</b>	<b>0.1779</b>
	FRCNN (RGB)	0.8344	0.7953	0.8144	0.2147

307 3-channel artificial RGB images achieved the best *recall* of 0.9572 for Dataset  
308 A. For dataset B, the *recall* of FRCNN marginally improved FCN-AlexNet  
309 but FCN-AlexNet produced more *FPIs*. Overall, FRCNN achieved the best  
310 *F1-Score* with 0.9321 and 0.9255 on Dataset A and Dataset B, respectively.

#### 311 4.1.2. Composite dataset analysis

312 When compared the composite results, FCN-AlexNet and FRCNN im-  
313 proved in terms of *recall* but with poorer performance in *precision*. These  
314 were due to the methods detecting more regions when trained on two datasets  
315 with different modalities. However, for FRCNN with the 3-channel artificial  
316 RGB technique, the results were less satisfactory for all the metrics. This  
317 has demonstrated that even though 3-channel artificial RGB images proved  
318 to improve the *recall* of Dataset A, which can be caused by introduction of

Table 3: Comparison of performance metrics based on *IoU* for ROI detection in BUS dataset. FRCNN is Faster-RCNN Inception-ResNet-v2 on concatenated grayscale BUS images whereas FRCNN (RGB) is Faster-RCNN Inception-ResNet-v2 on 3-channel artificial RGB BUS images. FCN-AlexNet represents transfer learning FCN-AlexNet. Bold indicates the best result for each category and underline indicates the best result for the Dataset. *STD* represents standard deviation.

Dataset	Method	<i>IoU</i> ( <i>mean</i> $\pm$ <i>STD</i> )	<i>Recall</i>	<i>Precision</i>	<i>F1-Score</i>	<i>FPI</i>
A	FCN-AlexNet	0.7800 $\pm$ 0.1069	0.8624	0.7684	0.8127	0.2778
	FRCNN	0.8447 $\pm$ 0.0946	0.8838	<b><u>0.9003</u></b>	0.8920	<b><u>0.1046</u></b>
	FRCNN (RGB)	<b><u>0.8535<math>\pm</math>0.0888</u></b>	<b><u>0.9358</u></b>	0.8818	<b><u>0.9080</u></b>	0.1340
B	FCN-AlexNet	0.7145 $\pm$ 0.1123	0.6749	0.6395	0.6567	0.3804
	FRCNN	<b><u>0.8363<math>\pm</math>0.0863</u></b>	<b><u>0.8773</u></b>	<b><u>0.8994</u></b>	<b><u>0.8882</u></b>	<b><u>0.0982</u></b>
	FRCNN (RGB)	0.8254 $\pm$ 0.0919	0.8221	0.8481	0.8349	0.1472
(A+B) on A	FCN-AlexNet	0.7837 $\pm$ 0.1066	0.8716	0.7703	0.8178	0.2778
	FRCNN	0.8496 $\pm$ 0.0904	<b>0.9205</b>	<b>0.8600</b>	<b>0.8892</b>	<b>0.1601</b>
	FRCNN (RGB)	<b><u>0.8532<math>\pm</math>0.0860</u></b>	0.7584	0.7230	0.7403	0.3105
(A+B) on B	FCN-AlexNet	0.7537 $\pm$ 0.1151	0.7485	0.6354	0.6873	0.4295
	FRCNN	0.8395 $\pm$ 0.0930	<b><u>0.8896</u></b>	<b>0.7796</b>	<b>0.8310</b>	<b>0.2515</b>
	FRCNN (RGB)	<b><u>0.8399<math>\pm</math>0.0896</u></b>	0.7485	0.7135	0.7305	0.3006

noisy data and hence become less robust across datasets. Overall, FRCNN is the most robust method across different datasets.

Since the measurement solely based on the *detected point* of the bounding box could be misleading, the following section reports the results based on overlap criterion – *IoU*.

#### 4.2. Evaluation based on *IoU*

Table 3 summarises the results based on the overlap criterion of *IoU* greater than 0.5. We report the results based on single datasets and composite dataset.

##### 4.2.1. Within dataset analysis

Since the overlap criterion followed a more strict rule, we observed all the performance metrics were poorer when compared to the *detected point*. Particularly the performance of FCN-AlexNet notably decreased for all the



332 evaluation. Interestingly, the FRCNN with 3-channel artificial RGB im-  
 333 ages worked the best on Dataset A with *recall* of 0.9358 and *F1-Score* of  
 334 0.9080. However, FRCNN achieved the best results on Dataset B with *recall*  
 335 of 0.8773, *precision* of 0.8994 and *F1-Score* of 0.8882. We observed the FR-  
 336 CNN with 3-channel artificial RGB images has achieved the best *IoU* when  
 337 evaluated on Dataset A.

#### 338 4.2.2. Composite dataset analysis

339 Similar to the results on *detected point*, FRCNN was the most robust  
 340 algorithm for the composite dataset analysis across all the performance met-  
 341 rics. FCN-AlexNet has shown marginal improvement when compared to the  
 342 within dataset analysis. However, FRCNN with 3-channel artificial RGB  
 343 images has deteriorated with very poor results. This has demonstrated that  
 344 even though 3-channel artificial RGB images proved to improve the *recall*  
 345 and *F1-Score* of Dataset A, it is not robust across datasets. A similar find-  
 346 ing shows FRCNN is more robust across different datasets when measured  
 347 by *IoU*. To further demonstrate the result, the following section reports qual-  
 348 itative analysis.

#### 349 4.3. Visual Comparison

350 Figure 7 visually compares the results of the proposed methods and FCN-  
 351 AlexNet. The yellow boxes indicate ground truth, the green boxes indicate  
 352 *TP* when *IoU* greater than 0.5, the red boxes indicate *FP*, the green ‘\*’  
 353 indicates *TP* for *detected point*, the red ‘\*’ indicate *FP* for *detected point*. The  
 354 first row of Figure 7 shows a best case for all the algorithms, the second row  
 355 shows the detected lesion by FRCNN but not FCN-AlexNet, and the third  
 356 row illustrates a complex case where all the algorithms achieved different  
 357 results. It is interesting to observe that FCN-AlexNet has a *TP* for *detected*  
 358 *point* but a *FP* for *IoU* criterion.

#### 359 4.4. Summary

360 From the results, we summarise our observations as follow:

- 361 • The overall performance of FCN-AlexNet was better on the composite  
 362 dataset. This implies that it is more suitable for larger heterogeneous  
 363 scale of dataset.

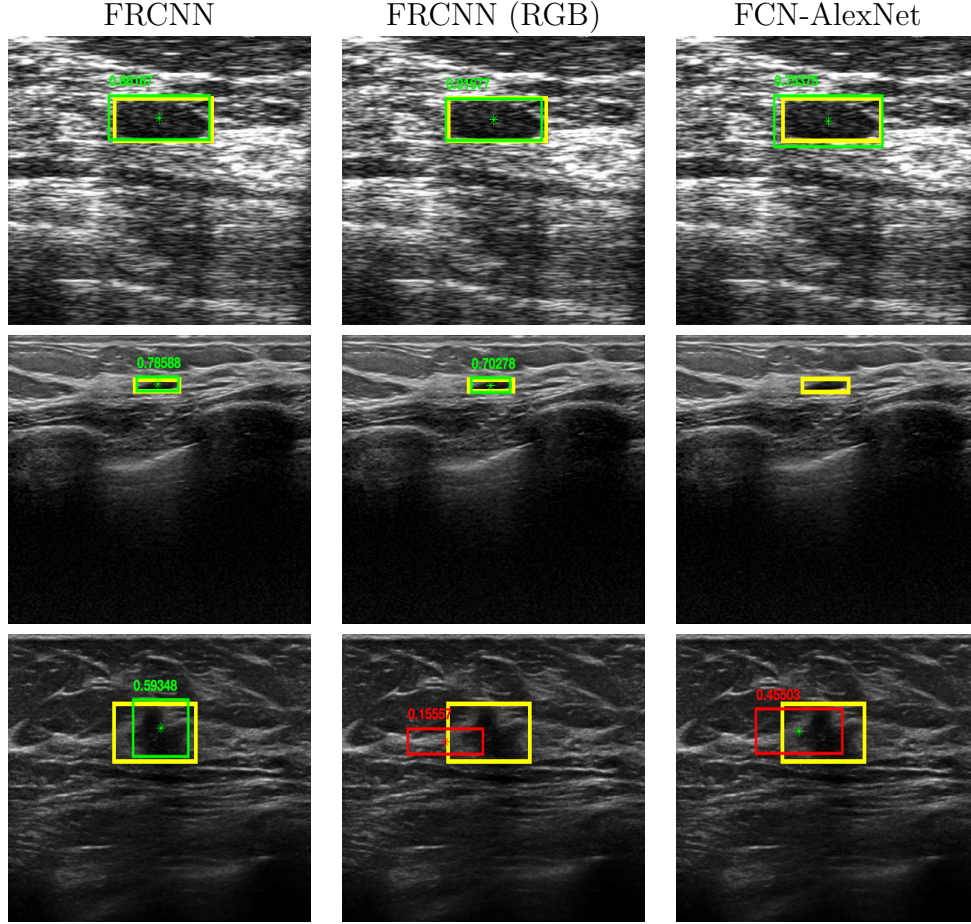


Figure 7: Examples cases from Dataset A and B to illustrate the performance of the lesion detection algorithms. The yellow rectangle indicates the ground truth, the “\*” is the *detected point*, green rectangle is the *TP* and red rectangle is the *FP*. The first row (image from Dataset A) shows an easy case where all methods detected the lesion. The second row (image from Dataset B) illustrate a case where the lesion is small and only detected by FRCNN (both with and without 3-channel artificial RGB images). The third row (image from Dataset B, based on the results of composite dataset analysis) shows an image with complex shadow and all the algorithms produced different results.

- The overall performance of FRCNN was better when assessed within individual dataset (see underlined results in Table 2 and Table 3). This is an indication that it is suitable for single source datasets.

- 367 • The proposed 3-channel artificial RGB method has potential to improve  
368 the *recall* but may not be suitable for images with different resolution.  
369 In our experiment, it only performed well on Dataset A, but not Dataset  
370 B. Current results are inconclusive and required further investigation.
- 371 • The overall results of FRCNN has a higher *mean IoU* and a lower  
372 *Standard Deviation* when compared to FCN-AlexNet.
- 373 • The limitation of this paper is the comparison of FCN-AlexNet with  
374 Faster R-CNN Inception-ResNet-v2, where the differences between the  
375 two networks could be overestimated. This potential bias is due to the  
376 two very different backbones used.

## 377 5. Conclusion

378 This paper proposed the use of the most accurate object detection deep  
379 learning framework – Faster-RCNN with Inception-ResNet-v2 – for breast  
380 ultrasound lesion detection and localisation. It investigated the use of a 3-  
381 channel artificial RGB technique, and the applicability to transfer learning  
382 in smaller datasets. Moreover, we showed that the Faster R-CNN approach  
383 obtains the best results compared to current state of the art when evaluated  
384 on two datasets using the *detected point* measurement and overlap criterion.  
385 These were then presented in four popular metrics: *recall*, *precision*, *F1-Score*  
386 and *FPI*.

387 The results showed Faster-RCNN with Inception-ResNet-v2 was the most  
388 robust algorithm across two datasets and worked well on small datasets.  
389 Although FCN-AlexNet achieved good results when evaluated with *detected*  
390 *point*, its performances deteriorated when evaluated using the intersection  
391 over union *IoU* as the criterion. In addition, the new 3-channel artificial RGB  
392 technique showed improved results when evaluated on Dataset A. However,  
393 the proposed 3-channel artificial RGB technique was not suitable for either  
394 Dataset B or the composite dataset. Further areas to improve our work  
395 include:

- 396 • Investigation in using different type of image manipulation techniques  
397 will have potential in improving the use of this 3-channel artificial RGB  
398 technique.

- 399 • To overcome the limitation of this paper, the use of a different feature  
400 extraction network, such as Feature Pyramid Network (FPN ResNet-  
401 101) should be investigated to evaluate the performance of the deep  
402 learning approach.
- 403 • Increase the volume of the datasets by data collection or introducing  
404 data-augmentation techniques such as albumentation (image augmen-  
405 tation and composition of image augmentation).

406 We demonstrated the use of state-of-the-art computer vision object de-  
407 tection algorithm on BUS lesion localisation. This is an important step  
408 forward to improve the lesion detection of BUS. We recommended the use of  
409 *IoU* (equivalent to Dice Coefficient Index, which is commonly used in lesion  
410 segmentation) in lesion detection as it is more reliable when compared to  
411 the *detected point*. Our work provides an important benchmark for future  
412 research.

## 413 Acknowledgment

414 The authors would like to thanks Prapavesis et al. for the permission to  
415 use Dataset A.

## 416 Reference

- 417 [1] H. Cheng, J. Shan, W. Ju, Y. Guo, L. Zhang, Automated breast cancer  
418 detection and classification using ultrasound images: A survey, Pattern  
419 Recognition 43 (1) (2010) 299 – 317.
- 420 [2] [link].  
421 URL <http://breastcancernow.org>
- 422 [3] W. Berg, L. Gutierrez, M. NessAiver, W. B. Carter, M. Bhargavan,  
423 R. Lewis, O. Ioffe, Diagnostic accuracy of mammography, clinical exam-  
424 ination, us, and mr imaging in preoperative assessment of breast cancer,  
425 Radiology 233 (3) (2004) 830–849.
- 426 [4] K. M. Kelly, J. Dean, W. S. Comulada, S.-J. Lee, Breast cancer de-  
427 tection using automated whole breast ultrasound and mammography in  
428 radiographically dense breasts, European radiology 20 (3) (2010) 734–  
429 742.

- 430 [5] A. Stavros, C. Rapp, S. Parker, Breast Ultrasound, 1st Edition, 978-  
431 0397516247, LWW, 1995.
- 432 [6] B. Huynh, K. Drukker, M. Giger, Mo-de-207b-06: Computer-aided di-  
433 agnosis of breast ultrasound images using transfer learning from deep  
434 convolutional neural networks, Medical Physics 43 (6) (2016) 3705–3705.
- 435 [7] M. H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentís, R. Zwigelaar, A. K.  
436 Davison, R. Martí, Automated breast ultrasound lesions detection using  
437 convolutional neural networks, IEEE journal of biomedical and health  
438 informatics 22 (4) (2018) 1218–1226.
- 439 [8] M. H. Yap, E. Edirisinghe, H. Bez, Processed images in human percep-  
440 tion: A case study in ultrasound breast imaging, European Journal of  
441 Radiology 73 (3) (2010) 682–687.
- 442 [9] W. Gómez-Flores, B. A. Ruiz-Ortega, New fully automated method for  
443 segmentation of breast lesions on ultrasound based on texture analysis,  
444 Ultrasound in medicine & biology 42 (7) (2016) 1637–1650.
- 445 [10] M. H. Yap, E. A. Edirisinghe, H. E. Bez, A novel algorithm for initial  
446 lesion detection in ultrasound breast images, Journal of Applied Clinical  
447 Medical Physics 9 (4) (2008) 181–199.
- 448 [11] K. Drukker, M. L. Giger, C. J. Vyborny, E. B. Mendelson, Computerized  
449 detection and classification of cancer on breast ultrasound, Academic  
450 Radiology 11 (5) (2004) 526–535.
- 451 [12] J. Shan, H. Cheng, Y. Wang, Completely automated segmentation ap-  
452 proach for breast ultrasound images using multiple-domain features, Ul-  
453 trasound in Medicine and Biology 38 (2) (2012) 262–275.
- 454 [13] G. Pons, R. Martí, S. Ganau, M. Sentis, J. Martí, A feasibility study  
455 of lesion detection using deformable part model in breast ultrasound  
456 images, in: Iberian Conference on Pattern Recognition and Image Anal-  
457 ysis, Vol. 7887, 2013, pp. 269–276.
- 458 [14] M. Byra, M. Galperin, H. Ojeda-Fournier, L. Olson, M. O’Boyle,  
459 C. Comstock, M. Andre, Breast mass classification in sonography with  
460 transfer learning using a deep convolutional neural network and color  
461 conversion, Medical physics (2018).

- 462 [15] M. H. Yap, M. Goyal, F. M. Osman, R. Martí, E. Denton, A. Juetten,  
463 R. Zwiggelaar, Breast ultrasound lesions recognition: end-to-end deep  
464 learning approaches, *Journal of Medical Imaging* 11007 (2019) 1.
- 465 [16] M. H. Yap, E. Edirisinghe, H. Bez, Computer aided detection and recog-  
466 nition of lesions in ultrasound breast images, in: *Innovations in Data  
467 Methodologies and Computational Algorithms for Medical Applications*,  
468 IGI Global, 2012, pp. 125–152.
- 469 [17] C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, Inception-v4, inception-  
470 resnet and the impact of residual connections on learning., in: *AAAI*,  
471 Vol. 4, 2017, p. 12.
- 472 [18] Y.-L. Huang, D.-R. Chen, Y.-K. Liu, Breast cancer diagnosis using im-  
473 age retrieval for different ultrasonic systems, in: *Image Processing, 2004.  
474 ICIP'04. 2004 International Conference on*, Vol. 5, IEEE, 2004, pp. 2957–  
475 2960.
- 476 [19] F. M. Osman, M. H. Yap, The effect of filtering algorithms for breast  
477 ultrasound lesions segmentation, *Informatics in Medicine Unlocked* 12  
478 (2018) 14–20.
- 479 [20] M. A. Kupinski, M. L. Giger, Automated seeded lesion segmentation  
480 on digital mammograms, *IEEE Transactions on medical imaging* 17 (4)  
481 (1998) 510–517.
- 482 [21] K. Drukker, M. L. Giger, K. Horsch, M. A. Kupinski, C. J. Vyborny,  
483 E. B. Mendelson, Computerized lesion detection on breast ultrasound,  
484 *Medical Physics* 29 (7) (2002) 1438–1446.
- 485 [22] M. H. Yap, E. A. Edirisinghe, H. E. Bez, Object boundary detection in  
486 ultrasound images, in: *The 3rd Canadian Conference on Computer and  
487 Robot Vision (CRV'06)*, IEEE, 2006, pp. 53–53.
- 488 [23] B. Liu, H. Cheng, J. Huang, J. Tian, X. Tang, J. Liu, Fully automatic  
489 and segmentation-robust classification of breast tumors based on local  
490 texture analysis of ultrasound images, *Pattern Recognition* 43 (1) (2010)  
491 280–298.
- 492 [24] Y. Ikeda, D. Fukuoka, T. Hara, H. Fujita, E. Takada, T. Endo,  
493 T. Morita, Development of a fully automatic scheme for detection of

- 494 masses in whole breast ultrasound images, *Medical physics* 34 (11)  
495 (2007) 4378–4388.
- 496 [25] G. Pons, R. Martí, S. Ganau, M. Sentís, J. Martí, Computerized detec-  
497 tion of breast lesions using deformable part models in ultrasound images,  
498 *Ultrasound in Medicine & Biology* 40 (9) (2014) 2252–2264.
- 499 [26] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553)  
500 (2015) 436–444.
- 501 [27] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma,  
502 Z. Huang, A. Karpthy, A. Khosla, M. Bernstein, et al., Imagenet  
503 large scale visual recognition challenge, arXiv preprint arXiv:1409.0575  
504 (2014).
- 505 [28] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan,  
506 P. Dollár, C. L. Zitnick, Microsoft coco: Common objects in context, in:  
507 *European conference on computer vision*, Springer, 2014, pp. 740–755.
- 508 [29] M. Goyal, M. H. Yap, N. D. Reeves, S. Rajbhandari, J. Spragg, Fully  
509 convolutional networks for diabetic foot ulcer segmentation, in: *2017*  
510 *IEEE International Conference on Systems, Man, and Cybernetics*  
511 (SMC), IEEE, 2017, pp. 618–623.
- 512 [30] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fis-  
513 cher, Z. Wojna, Y. Song, S. Guadarrama, et al., Speed/accuracy trade-  
514 offs for modern convolutional object detectors, in: *IEEE CVPR*, Vol. 4,  
515 2017.
- 516 [31] M. Goyal, N. D. Reeves, S. Rajbhandari, M. H. Yap, Robust methods  
517 for real-time diabetic foot ulcer detection and localization on mobile  
518 devices, *IEEE journal of biomedical and health informatics* 23 (4) (2018)  
519 1730–1741.
- 520 [32] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time  
521 object detection with region proposal networks, in: *Advances in neural*  
522 *information processing systems*, 2015, pp. 91–99.
- 523 [33] C. L. Zitnick, P. Dollár, Edge boxes: Locating object proposals from  
524 edges, in: *European conference on computer vision*, Springer, 2014, pp.  
525 391–405.

- 526 [34] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick,  
527 S. Guadarrama, T. Darrell, Caffe: Convolutional architecture for fast  
528 feature embedding, in: Proceedings of the 22nd ACM international con-  
529 ference on Multimedia, ACM, 2014, pp. 675–678.
- 530 [35] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for se-  
531 mantic segmentation, in: Proceedings of the IEEE Conference on Com-  
532 puter Vision and Pattern Recognition, 2015, pp. 3431–3440.
- 533 [36] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for  
534 biomedical image segmentation, in: International Conference on Medical  
535 Image Computing and Computer-Assisted Intervention, Springer, 2015,  
536 pp. 234–241.
- 537 [37] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning  
538 applied to document recognition, Proceedings of the IEEE 86 (11) (1998)  
539 2278–2324.